....

Guidelines for Using Student Work for Judging Prompts

Rhode Island Skills Commission

....

Explanation and Considerations for Use

This document was developed by the Rhode Island Skills Commission. It describes a process for the revision of tasks, one of several steps in task development. After your school has piloted and scored new tasks you may want to engage in this process of using student work to evaluate and revise common tasks. It is important to evaluate common tasks for validity and reliability on a regular basis. This document represents one approach to this process; your school may choose to adopt it or may want to explore other approaches.

http://www.ride.ri.gov/highschoolreform/dslat/
October, 2005

Guidelines for Using Student Work to Revise Tasks

One of the most effective ways to see how well a rubric works is to use it to assess student work. When a group—even a small group—of scorers uses the calibration protocol to score a task, all the ambiguities and shortcomings of the rubric stand out in high relief. This is because every time the scoring team disagrees on the way they are scoring the task they appeal to the rubric to settle the disagreement. And very often, the rubric is inadequate to the task. When scorers search the rubric for the subtleties that would settle their disagreement, they often find that it is not there.

Experience has shown that this is not a problem that people who write rubrics can entirely avoid. Instead, the reason this problem keeps coming up is because of the interaction between scorers and the rubric and the relation of the rubric to the standards the task assesses and the task prompt. Because many people will use a rubric, it is impossible to predict how all of them will interpret the words that actually make up the rubric. However, using the rubric at least once before finalizing it for use will point out almost all of the most obvious errors and go a very long way to making the rubric clear and usable.

There are some basic guidelines that help make a draft rubric better. They are:

Make sure that the criteria for performance describe as clearly as possible what students have to DO to demonstrate proficiency. For example, for the expectation "students can create an organizing structure", a description of what students DO to create that organizing structure might include that they 1) introduce the problem or issue, 2) lay out the elements of the problem or issue in a logically clear sequence, 3) consider alternative solutions to the problem or approaches to the issue, and 4) advocate for a "best" solution or a summary judgment that brings closure to all the issues raised. Here, the important point is to think through ALL the things a student must do to create an organizing structure and then describe them in the rubric.

Be as clear as possible when you quantify criteria. Often performances are distinguished by whether, and how often things are included or not. This creates different kinds of lists. For example, a list that includes **every** element as a criterion would be "uses sensory detail, more than one voice, and uses at least two perspectives". This list is different from a list that only requires the use of one of these elements, as in "uses sensory detail, **or** more than one voice, **or** more than one perspective". It is also different from a list that requires one element and one of the last two elements, as in "must use sensory detail and uses more than one voice or more than one perspective". The point is that the rubric(s) writer needs to be very clear how the criteria will count in scoring a performance and to describe that with as little room for interpretation as possible.

Create parallelism across performance levels. Each level (exemplary, meets, nearly meets, below, little or no evidence) of an expectation requires a distinct description in a rubric. Each description should use the language of every other description insofar as

possible. This creates a "core" description, which should be built around meeting the standard. Having a core description makes it possible to create differences across levels by adding, omitting, or changing parts of the core description. This is an example of changing a core description to create the criteria for a different performance level.

Here, the core statement listing the organizational elements required to meet standard was revised to describe a performance that nearly meets the standard by saying "however, one or more of these elements is not fully developed".

Core Description for Meeting Standard	Revised Core Description for Nearly Meeting Standard
The reflection is clearly organized from beginning to end. This includes an effective opening, body, and closure.	The reflection includes an opening, body, and closure; however, one or more of these elements is not fully developed.

Use clear descriptions of qualities. Sometimes rubrics resort to words such as "masterful" or "adequate" or "poorly" to distinguish levels of performance. This leaves scorers to guess what the words might possibly mean and to engage in inevitable arguments about how they apply to any given task. Often these words get used because not enough thought has been given about what they are trying to describe. For example, a "masterful" performance might be a performance in which the student "organizes the introduction in novel and appropriate way", or "repeatedly uses original language appropriately", or "creates a high level of suspense by presenting evidence in a counter-intuitive way". Any one of these phrases gives more detail about what students need to do than the phrase "masterful", and they all put the scorers in a much stronger position to decide whether the performance is indeed masterful.

Use single criteria. Sometimes rubrics contain two (or more) criteria, which create possible dilemmas. For example, what would happen in the rubric shown below if a scorer using the rubric below scored a task as meeting standard on the part of the criteria referring to clearly identifying a condition, but as nearly meeting that part of the criteria that conveys personal insight? While this is not an entirely unmanageable situation, it is much simpler to write a rubric that has only one criterion per row.

Meets Standard	Nearly Meets Standard
Clearly identifies a condition, a situation, or an issue that addresses the prompt. --------------------------------------- Analysis conveys personal insight into significance of the condition, situation, or issue.	Identifies a condition, a situation, or an issue but does not adequately address the prompt. --------------------------------------- Attempts to convey personal insight.